# Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments

Ning Wang[1], Ya Yang[2], Michael J. Moore[3], Samuel F. Brockington[4], Joseph F. Walker[1], Joseph W Brown[5], Bin Liang[1], Tao Feng[4], Caroline Edwards[3], Jessica Mikenas[3], Julia Olivieri[3], Vera Hutchison[3], Alfonso Timoneda[4], Tommy Stoughton[6], Raúl Puente[7], Lucas C. Majure[7,8], Urs Eggli[9], and Stephen A. Smith[1]

[1]Department of Ecology & Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048, USA

[2]Department of Plant and Microbial Biology, University of Minnesota-Twin Cities. 1445 Gortner Avenue, St. Paul, MN 55108 USA

[3]Department of Biology, Oberlin College, Science Center K111, 119 Woodland St., Oberlin, Ohio 44074 USA

[4]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[5]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

[6]Center for the Environment, MSC 63, Plymouth State University, 17 High Street Plymouth, NH 03264 USA

[7]Department of Research, Conservation and Collections, Desert Botanical Garden, 1201 N. Galvin Pkwy, Phoenix, AZ 85008 USA

[8]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611 USA

[9]Sukkulenten-Sammlung Zürich, Mythenquai 88, CH-8002 Zürich, Switzerland

Corresponding authors: Ning Wang, Email: ningwum@umich.edu

Stephen A. Smith, E-mail: eebsmith@umich.edu

Abstract:

Several plant lineages have evolved adaptations that allow survival in extreme and harsh environments including many within the plant clade Portulacineae (Caryophyllales) such as the Cactaceae, Didiereaceae of Madagascar, and high altitude Montiaceae. Here, using newly generated transcriptomic data, we reconstructed the phylogeny of Portulacineae and examine potential correlates between molecular evolution within this clade and adaptation to harsh environments. Our phylogenetic results were largely congruent with previous analyses, but we identified several early diverging nodes characterized by extensive gene tree conflict. For particularly contentious nodes, we presented detailed information about the phylogenetic signal for alternative relationships. We also analyzed the frequency of gene duplications, confirmed previously identified whole genome duplications (WGD), and identified a previously unidentified WGD event within the Didiereaceae. We found that the WGD events were typically associated with shifts in climatic niche and did not find a direct association with WGDs and diversification rate shifts. Diversification shifts occurred within the Portulacaceae, Cactaceae, and Anacampserotaceae, and while these did not experience WGDs, the Cactaceae experienced extensive gene duplications. We examined gene family expansion and molecular evolutionary patterns with a focus on genes associated with environmental stress responses and found evidence for significant gene family expansion in genes with stress adaptation and clades found in extreme environments. These results provide important directions for further and deeper examination of the potential links between molecular evolutionary patterns and adaptation to harsh environments.

key words: gene tree conflict, gene duplication, molecular evolution, paleopolyploidy, Portulacineae, stress adaptation

## Introduction

Temperature and water availability are two major ecological determinants of plant distribution and survival (e.g., Peel et al. 2007). Plants living in harsh environments such as low precipitation, extreme temperature ranges, intense sun and/or dry winds, often exhibit specialized morphological and physiological adaptations to abiotic stresses (Gibson 1996; Kreps et al. 2002). While the morphological adaptations are perhaps easier to identify, these long-term selective pressures are expected to leave genetic and genomic signatures in lineages adapted to extreme environments. Researchers have addressed these questions in several ways, as for example exploring the gene regions and functions associated with adaptation in certain species or small groups (e.g., Christin et al. 2007).

Over the last decade, genomic data have increased dramatically and provide new ways to address macroevolutionary questions. For example, transcrimptomic and genomic data sets provide the necessary means to more accurately identify whole genome duplication events (WGD, e.g., Jiao et al. 2011; Yang et al. 2015, Parks et al. 2018). WGDs have been identified in major clades throughout plants and have previously been suggested to be associated with adaptations to extreme environments (e.g., Stebbins 1971; Soltis and Soltis 2000; Brochmann et al. 2004), speciation/diversification (Stebbins 1971; Wood et al. 2009), and success at colonizing new regions (Soltis and Soltis 2000). We continue to identify more WGDs as we increase taxon sampling (Yang et al. 2018). Broadly sampled genomic data also facilitate analyses of lineage specific molecular evolution such as gene family expansion and differential selection among genes and taxa. Finally, genomic data allow for more thorough and accurate phylogenetic reconstruction. In particular, recent studies have illustrated that phylogenetic conflict is common, but the impact of this conflict on phylogenetic reconstruction varies across the tree of life due to different macroevolutionary processes (e.g., Shen et al. 2017; Walker et al. 2017). Genomic data allow us to examine conflict and identify the processes that contribute.

The Portulacineae comprise ~ 2200 species in nine families and 160 genera (Nyffeler and Eggli 2010a; Angiosperm Phylogeny Group 2016) that are inferred to have diverged from its sister group, the Molluginaceae, ~54 Mya (Arakaki et al. 2011). Species in this suborder are incredibly morphologically diverse, ranging from annual herbs to long-lived stem succulents to trees, and include several iconic plants such as the cacti and Malagasy Didiereaceae. They also exhibit variable habitat preferences ranging from rainforests to deserts (Smith et al. 2018a). While some species are distributed worldwide, the majority are restricted to seasonally dry areas of the northern and southern hemisphere, under either hot arid or cold arid conditions (Hernández-Ledesma et al. 2015). Many specialized traits such as fleshy or succulent stems, leaves, and/or underground perennating structures, have arisen in association with the adaptation of this clade to xeric, alpine and arctic environments. Portulacineae also include

84 several transitions to Crassulacean acid metabolism (CAM) and $C_4$ photosynthesis that improve
85 photosynthetic efficiency and hence minimize water loss in hot and dry environments compared to $C_3$
86 photosynthesis (Edwards and Walker 1983; Borland et al. 2009; Edwards and Ogburn 2012).
87 Several clades within the Portulacineae have adapted to a wide array of harsh environments.
88 The Cactaceae encompass roughly 80% of species in Portulacineae (~1850 species; Nyffeler and Eggli
89 2010b) and represent perhaps the most spectacular radiation of succulent plants (Mauseth 2006;
90 Arakaki et al. 2011). In addition to succulent structures that enable water storage, Cactaceae have an
91 array of adaptations to cope with arid and semiarid conditions. For example, many have spines that
92 reduce air flow on surface area for evapotranspiration, and deter herbivores (Anderson 2001). They
93 also have extensive, but relatively shallow, root systems that facilitate quick absorption of rainfall
94 (Gibson and Nobel 1990). The Didiereaceae also exhibit specialized adaptations, including thorns and
95 succulent leaves, to the hot and seasonally dry environments of Madagascar. Unlike the Cactaceae and
96 Didiereaceae that are typically found in tropical and subtropical areas, the Montiaceae have a more
97 cosmopolitan distribution with most species occurring in colder environments. Some species of
98 Montiaceae, especially in the genera *Claytonia* and *Montia,* are adapted to high alpine zones and/or the
99 high Arctic (Ogburn and Edwards 2015; Stoughton et al. 2017). Repeated evolution into harsh
100 environments suggests the possibility that the clade as a whole may be predisposed to adaptation to
101 extreme environments.
102 Recent phylogenetic work has helped to resolve many relationships within the Portulacineae,
103 but the earliest diverging relationships are only moderately or poorly supported (Arakaki et al. 2011;
104 Ogburn and Edwards 2015; Moore et al. 2017; Yang et al. 2018; Walker et al. 2018a). This may reflect
105 a rapid radiation, uncertainty due to gene tree conflict or whole genome duplications, or other
106 macroevolutionary processes. Recent studies have inferred several whole genome duplication (WGD)
107 events (i.e., paleopolyploidy) within the Portulacineae, including one at the base of the clade (Yang et
108 al. 2018). However, Arakaki et al. (2011) primarily used chloroplast loci and Yang et al. (2018)
109 employed transcriptomes, but taxon sampling within Portulacineae was limited. Increased
110 transcriptomic sampling may allow for the detection of additional WGD events.
111 In addition to phylogenetic analyses, genomic-scale data also facilitates extensive molecular
112 evolution analyses. In particular, these data can provide preliminary, but essential, information on the
113 genetic basis of adaptation to abiotic stresses through gene duplication and other evolutionary
114 processes. For example, genes that have been identified to function in response to drought, cold, and
115 heat conditions such as those involved in the abscisic acid (ABA) signaling pathway [e.g., ABRE-
116 binding protein/ABRE-binding factors, the dehydration-responsive element binding (DREB) factors,

117 and the NAC transcription factors (e.g., Qin et al. 2004, Nakashima et al. 2012)] provide good
118 candidates for detecting common signatures in plant lineages, such as Portulacineae, with stress
119 adaptation. Of course, plant stress responses can be complicated by the occurrence of multiple stresses
120 (Nakashima et al. 2014), and how genes mediate stress responses in most plant taxa are still poorly
121 known. Nevertheless, molecular analyses of potential stress related genes provide essential information
122 for future targeted studies. Because the Portulacineae include many species that occupy high stress
123 environments, they represent an excellent clade with which to examine the genetic basis of adaptation
124 to abiotic stresses. By examining transcriptomes, we can identify lineage-specific expansions and/or
125 positive selection among taxa that occupy these environments and gain new insights into the evolution
126 of stress responses that may be broadly applicable to plants.
127 In this study, we analyzed 82 transcriptomes, 47 of which were newly generated, to better
128 understand the evolutionary history of the Portulacineae. By closely examining patterns of gene
129 duplication and paleopolyploidy, gene-tree/species-tree conflict, and clade-specific evolutionary
130 patterns of genes associated with stress-related responses, we aimed to 1) identify additional WGDs; 2)
131 examine gene tree conflict as a source of uncertainty in the reconstruction of relationships within the
132 Portulacineae; 3) estimate phylogenetic signal for particularly contentious relationships; and 4)
133 examine whether gene/genome duplications were associated with clade-specific adaptive traits and/or
134 plant diversification rate.
135
136 **Results and Discussion**
137 **Phylogeny of Portulacineae**
138 Our concatenated supermatrix contained 841 gene regions and had a total aligned length of 1,239,871
139 bp. Gene and character occupancy were 95.4% and 84.5%, respectively. Of the ingroup taxa, only
140 *Pereskia aculeata* had a gene occupancy less than 80%, while the majority (67 taxa) had over 90% of
141 genes present in the final supermatrix. Both the concatenated maximum likelihood tree (hereafter the
142 CML tree) and the maximum quartet support species tree (MQSST) recovered the same topology with
143 similar support values (fig. 1), all of which were largely consistent with previous estimates of the
144 Portulacineae phylogeny (e.g., Arakaki et al. 2011; Moore et al. 2017) despite different datasets and
145 datatypes. For example, our analysis recovered the ACPT clade (Anacampserotaceae, Cactaceae,
146 Portulacaceae and Talinaceae; Nyffeler 2007; Nyffeler and Eggli 2010b; Ocampo and Columbus 2010;
147 Arakaki et al. 2011; Moore et al. 2017), with high support across genes trees for a topology of (((A, P),
148 C), T) (figs. 1-2) that generally agrees with that of Moore et al. (2017) using a targeted enrichment
149 approach. However, because bootstrap values may be a poor indicator of support in large

150 phylogenomic datasets (e.g., erroneous topologies can be increasingly supported as more sequence data
151 are added, Alfaro et al. 2003; Phillips et al. 2004), we also conducted gene tree conflict analyses (fig. 2,
152 Smith SA et al. 2015). For example, while we found strong support (100% bootstrap for both MQSST
153 and CML trees) for the sister relationship of Molluginaceae and Portulacineae (fig. 1), consistent with
154 several recent studies (Edwards and Ogburn 2012; Yang et al. 2015, 2018; Moore et al. 2017), gene
155 tree analyses highlighted that most genes were uninformative for or conflicted with this resolution. We
156 discuss the ACPT resolution in more detail below.
157 Our broad taxon sampling also allowed for resolution of relationships among early-diverging
158 lineages within Portulacineae families. We found support for the sister relationship of Didiereaceae
159 subfamilies Portulacarioideae (*Ceraria* and *Portulacaria*) and Didiereoideae (Madagascar Didiereaceae;
160 Applequist and Wallace 2003; fig. 1). Although many gene trees were uninformative at this node, the
161 ML topology was also the most frequent found among the gene trees (fig. 2). Relationships among the
162 genera within the Didiereoideae have been difficult to resolve with targeted-gene analyses (Applequist
163 and Wallace 2000; Nyffeler and Eggli 2010b). Here, we recovered strong support (100%) in gene tree
164 analyses for a clade including only *Alluaudia* and *Alluaudiopsis* (fig. 1). This result differs from
165 Bruyns et al. (2014) that found *Alluaudiopsis* to be sister to the remaining Didiereoideae.
166 Within Cactaceae, we recovered three major clades (*Pereskia* s.s., *Leuenbergeria,* and the core
167 cacti, as in Edwards et al. 2005 and Bárcenas et al. 2011) with strong support (> 97%, fig. 1), but with
168 substantial gene tree discordance (see below, fig. 2). Most genes (>85%) were uninformative or in
169 conflict (fig. 2) among the earliest diverging nodes of Cactaceae. Within the core cacti, we recovered
170 *Maihuenia* as sister to Cactoideae with strong support, and this clade as sister to Opuntioideae (fig. 1)
171 as found by Edwards et al. (2005) and Moore et al. (2017). Nevertheless, the position of *Maihuenia* has
172 been found to be highly unstable (e.g., within Opuntioideae, Butterworth and Wallace 2005; sister to
173 Opuntioideae, Nyffeler 2002; or sister to Opuntioideae+Cactoideae, Hernández-Hernández et al. 2014;
174 Moore et al. 2017). Unlike other areas within the Cactaceae that have high gene tree discordance, the
175 monophyly of Opuntioideae (e.g., Barthlott and Hunt 1993; Wallace and Dickie 2002; Griffith and
176 Porter 2009; Hernández-Hernández et al. 2011) and Cactoideae (Nyffeler 2002; Bárcenas et al. 2011)
177 had relatively low conflict (95-97% gene trees support these monophyletic groups; fig. 2). Within
178 Opuntioideae, however, there was a high level of discordance among gene trees. However, these
179 conflicts may have been influenced by limited taxon sampling as we only have one to six species
180 representing each tribe (figs. 1 and 2). The topology of the Cactoideae was largely congruent with
181 previous studies (e.g., Butterworth et al. 2002; Nyffeler 2002; Bárcenas et al. 2011; Hernández-
182 Hernández et al. 2011). Within Cacteae, the relationships among the five sampled genera were strongly

183 supported and are consistent with relationships recovered by Hernández-Hernández et al. (2011) and
184 Vázquez-Sánchez et al. (2013) using five loci. However, the core Cactoideae (sister to Cacteae,
185 Hernández-Hernández et al. 2011) were poorly supported in both the CML tree (52%) and the MQSST
186 inference (54.5%, fig. 1). With limited taxon sampling, some relationships (e.g., the monophyly of the
187 Pachycereeae) are largely consistent with previous studies (e.g., Nyffeler and Eggli 2010b), whereas
188 others (e.g., the position of *Gymnocalycium*) are still incongruent among studies (Arakaki et al. 2011;
189 Bárcenas et al. 2011; Hernández-Hernández et al. 2014).
190
191 **Assessing conflict at specific nodes**
192 Incongruence can result from several biological processes, including incomplete lineage sorting (ILS),
193 hybridization, and horizontal gene transfer (HGT). As more genomes and transcriptomes have become
194 available, our ability to analyze phylogenetic conflict and concordance across hundreds or thousands of
195 genes has increased significantly. We found that gene tree conflicts are prevalent within Portulacineae.
196 Several areas with high conflict occurred immediately after inferred WGD events (fig. 2) suggesting a
197 major role of gene duplication and loss.
198 While incorporating discordance into species tree analyses is important (Liu and Pearl 2007;
199 Liu et al. 2008), a close examination of gene-specific discordance as it pertains to support, or lack
200 thereof, for different clades is also valuable. Recently, studies have also demonstrated that only a few
201 genes in phylogenomic datasets can drive the resolution of nodes (e.g., Brown and Thomson 2017;
202 Shen et al. 2017; Walker et al. 2018b). We examined individual genes and their support for specific
203 phylogenetic relationships that had previously been identified to have significant conflicts including the
204 earliest divergences of Cactaceae (particularly the relationships among *Pereskia* s.s., *Leuenbergeria*,
205 and the core cacti; Edwards et al. 2005; Moore et al. 2017), the relationships among Portulacaceae,
206 Cactaceae, and Anacampserotaceae, and the positions of Basellaceae and Didiereaceae (e.g., Moore et
207 al. 2017, Walker et al. 2018a). This focused analysis allowed us to isolate the signal at particular nodes
208 without having to accommodate the many diverse processes that may shape conflict in other parts of
209 the tree. To address the early diverging Cactaceae, we examined the support for (*Leuenbergeria*,
210 (*Pereskia* + core cacti)) versus ((*Leuenbergeria* + *Pereskia*), core cacti) (fig. 3). 472 genes (385 out of
211 which have $\Delta \ln L > 2$) supported the grade, with 368 (287 out of which have $\Delta \ln L > 2$) supporting
212 *Leuenbergeria* + *Pereskia*. We also found that one gene (cluster4707, with homology to the "NF-X1-
213 type zinc finger protein") strongly supported *Pereskia* + core cacti (>110 ln *L* units), although with no
214 lineage-specific positive selection, gene duplication, or any obvious alignment problems. Overall, these
215 data support the resolution of (*Leuenbergeria*, (*Pereskia* + core cacti)).

The relationship of Portulacaceae, Anacampserotaceae, and Cactaceae have also been contentious. Portulacaceae (P) have been recovered as sister to Anacampserotaceae (A, Ogburn and Edwards 2015; Walker et al. 2018a; and recovered here), to Cactaceae (C, Arakaki et al. 2011), and to both (Nyffeler 2007, Nyffeler and Eggli 2010). We found that 476 of 841 genes supported A+P (343 with $\Delta \ln L > 2$), 164 genes supported C+P (91 with $\Delta \ln L > 2$), and 201 genes supported C+A (118 with $\Delta \ln L > 2$). While several processes may contribute to this conflict, the distribution of the alternative placements suggest ILS to be a source of conflict (as in Moore et al. 2017) though missing data complicates our ability to test this pattern. Gene tree conflict analyses, when requiring bootstrap support > 70%, confirm this as 276 genes were congruent with the species tree, 68 were congruent with C+P, and 85 were congruent with C+A (fig. 3A). Extensive gene duplications can confound phylogenetic reconstruction due to “incomplete sorting” of paralogs (see Moore et al. 2017). However, we find low level of gene duplications at the origin of PAC (1.3%, fig. 1). We found several outlying genes including one (cluster4488, homologous to the "ARABILLO 1-like" in *Beta vulgaris*) that supported C+P (>60 ln *L* units) and one (cluster7144, homologous to the "UV-B induced protein chloroplastic-like" gene in *Chenopodium quinoa*) that supported C+A strongly (>100 ln *L* units). We did not detect any obvious errors in orthology inference (e.g., Brown and Thomson 2017), alignment, or lineage-specific positive selection.

We conducted similar analyses for the resolution of Basellaceae and Didiereaceae (fig. 3B). 506 genes (337 with $\Delta \ln L > 2$) supported (Basellaceae, (Didiereaceae+ACPT)) and 334 genes (185 with $\Delta \ln L > 2$) supported Basellaceae+Didiereaceae as found by Soltis et al. (2011) and Anton et al. (2014). No significant outlying genes (the $\Delta \ln L$ ranging from -31.41 to 42.98 smoothly) were observed supporting either topology (fig. 3B).

All of the tests discussed demonstrated significant conflict among the gene regions. To address whether strong signal for hybridization contributed to the conflicting signal, we conducted hybridization analyses. The results from these analyses suggest that we cannot rule out ancient hybridization (table S1 and fig. S1). However, there is no other evidence (such as species history and geographical distribution) for hybridization between these lineages, and so we consider this to be further evidence of complex gene family evolution within the clade. Nevertheless, future analyses of synteny or other evidence may provide more direct evidence for or against hybridization as a source of conflict.

**Multiple whole-genome duplication events in Portulacineae**

248 Whole-genome duplications (WGD) have had a profound influence on the evolutionary history of
249 plants (e.g., Cui et al. 2006; Soltis and Soltis 2009; Jiao et al. 2011; Wendel 2015; Yang et al. 2015;
250 Smith et al. 2018a). Previous analyses have identified WGD events within the Caryophyllales (Yang et
251 al. 2015, 2018; Walker et al. 2017) including one in the ancestor of Portulacineae and two within
252 Portulacineae: in the lineage of *Basella alba*, and within Montiaceae. Improved taxon sampling enabled
253 us to confirm these events and infer additional putative WGD events as well as several instances of
254 large-scale gene duplications (fig. 1). Ks plots have been widely applied to infer WGDs but are not
255 without limitations (Parks et al. 2018). Higher Ks values (e.g., >2) are associated with increasingly
256 large error (Vanneste et al. 2013) and smaller Ks values (e.g., < 0.25) can be difficult to interpret due to
257 the errors in assembly, splice variants, or other anomalies. We used both gene tree and Ks analyses to
258 identify WGDs more confidently (fig. 1).
259 We found evidence for a previously identified WGD event in the ancestor of Portulacineae
260 (Yang et al. 2015, 2018) given the high percentage of gene duplications (13.4%) and a Ks peak (~0.4-
261 1.0) shared by all members in this clade. However, we did not detect a similar pattern of gene
262 duplications at the ancestor of Portulacineae+Molluginaceae (1%), confirming that Yang et al. (2018)'s
263 estimation could result from phylogenetic uncertainty. We found a high percentage of gene
264 duplications at the base of the Didiereoideae (50.1%) accompanied by a very recent Ks peak (between
265 0.05-0.10) for all the species within this clade. We note that the Didiereoideae is a relatively slow
266 evolving lineage with a recent Ks peak. Therefore, we rely heavily on corroborating both the Ks
267 analyses and gene tree analyses in order to confidently identify the WGD (i.e., Ks peak occurred < 0.1,
268 fig. S2). We also identified genome duplications within the Montiaceae. Yang et al. (2018) identified
269 WGD in the ancestor of *Claytonia* species and, with expanded sampling, we placed this WGD event at
270 the ancestor of *Lewisia* and *Claytonia* as all four species in this clade exhibited Ks peaks between 0.2–
271 0.4 (figs. S2 and S3). We also identified a high number of gene duplications in the two *Calyptridium*
272 species, accompanied by a very recent Ks peak (between 0.05–0.10; fig. 1).
273 While we identified many gene duplications (7.3%) at the origin of the Cactaceae, this was not
274 corroborated by Ks analyses. This is consistent with other recent studies (Walker et al. 2017; Yang et al.
275 2018). Several other nodes (e.g., nodes 50 and 39) exhibited similar patterns with high numbers of
276 duplications (3.8-6.4%) but no corresponding Ks peak. It is unclear what events may be associated with
277 these duplication events or whether gene loss, gene tree conflict, life history shifts, etc., may be
278 obscuring the detection of gene/genome duplications, but further investigation will help shed light on
279 these patterns and processes.
280 Although gene-tree-based methods can help identify the phylogenetic location of WGD events,

281 they are unable to infer WGD events that have occurred on terminal branches. To identify these, we
282 examined Ks peaks that are unique to one taxon and absent from close relatives. We inferred two such
283 cases: on the branch leading to *Basella alba* (Ks=0.45) and the branch leading to *Mollugo verticillata*
284 (Ks=0.25; fig. S2). Both species also exhibit higher chromosome numbers relative to close allies and
285 are consistent with a previous study (Yang et al. 2018).
286
287 **WGD, diversification, and climatic niche shifts**
288 Smith et al. (2018a) examined the potential associations between WGDs, diversification rate shifts, and
289 climatic niche shifts within the Caryophyllales and found that some WGD events were associated with
290 shifts in climatic niche but not diversification rate shifts. Here, we conducted similar analyses with a
291 focus on the Portulacineae. Specifically, we examined mean annual temperature, mean annual
292 precipitation, and the first axis of a principal component analysis on bioclimatic variables 1-19 (PCA1)
293 among clades with WGDs (Portulacineae, Didiereoideae, and Montiaceae), as well as the Cactaceae
294 using a species-level phylogeny. The Montiaceae, except for *Phemeranthus*, were treated together as
295 the species-level tree conflicted with the transcriptomic tree regarding internal resolution of clades
296 within the Montiaceae. Consistent with Smith et al. (2018a), WGD events were associated with shifts
297 in climate tolerance (fig. 4A and 4B). Montiaceae was associated with movement into colder
298 environments, as found in Smith et al. (2018a); Didiereoideae occupied a wetter climate than the sister
299 clade but with both clades occupying relatively seasonally dry areas (fig. S4). The Cactaceae exhibited
300 more variable rates of climatic niche evolution (fig. 4B), occupied a slightly hotter environment, but
301 did not exhibit a strong pattern of different climate occupancy. This does not negate climate expansion
302 within the clade, but suggests it may not be experienced by the entire clade.
303 While WGDs may not always be associated with shifts in niche occupancy, their association
304 appears to be common within the Portulacineae and the broader Caryophyllales. The Cactaceae did not
305 exhibit a WGD but instead extensive gene duplication and the climatic niche results for that clade are
306 less clear. Some have suggested that WGDs may be associated with diversification rate increases
307 (Landis et al. 2018). We examined this question on both the species-level phylogeny and a reduced
308 phylogeny of major lineages using the estimated diversity. In the species level phylogeny, we found
309 diversification rate shifts within Portulacaceae, Anacampserotaceae, and Cactaceae (fig. 4). In the
310 reduced phylogeny, we found shifts at the origin of core Cactaceae (fig. S5). The differences between
311 these are the result of the species level phylogeny being able to better place shifts while the reduced
312 phylogeny better able to represent to total diversity. Nevertheless, in either case we did not find a
313 correspondence between diversification shifts and WGDs.

314

315 **Gene families show broad expansions across Portulacineae**

316 Examining gene family expansion in depth has been instrumental in understanding functional trait
317 evolution in Caryophyllales. For example, gene expansion and neo-functionalization have been
318 implicated in the evolution of betalain pigmentation in Caryophyllales (Yang et al. 2015; Brockington
319 et al. 2015). Here, we explored gene family expansion in Portulacineae with respect to stress tolerance.
320 The top 20 most expanded gene families in Portulacineae (table 1) included genes encoding
321 transporters, proteases, cytoskeletal proteins and enzymes that are involved in photosynthetic pathways.
322 Several genes were associated with responses to drought: the gene encoding plasma membrane intrinsic
323 protein (PIP) can regulate the transport of water through membranes (Vandeleur et al. 2009), heat
324 shock proteins (Kiang and Tsokos 1998), and ubiquitin (Hochstrasser 2009). But perhaps the most
325 interesting gene expansions involved those encoding phosphoenol-pyruvate carboxylase (PEPC,
326 Silvera et al. 2014) and NADP-dependent malic enzyme catalyses (Ferreyra et al. 2003), which are key
327 enzymes involved in CAM/$C_4$ photosynthesis (Smith and Winter 1996; Cushman 2001). Lineage-
328 specific selection analyses on these two genes revealed that 10-13 internal branches are likely under
329 positive selection, with the most occurred within Cactaceae. The role of gene duplication in the
330 evolution of $C_4$ photosynthesis has been contentious, and some authors have proposed that neo-
331 functionalization of genes following duplication has not played a major role in the evolution of $C_4$
332 photosynthesis (e.g., Williams et al. 2012; van den Bergh et al. 2014). However, recent analyses
333 provide evidence that duplication and subsequent retention of genes are associated with the evolution
334 of $C_4$ photosynthesis (Emms et al. 2016), specifically including the NADP-dependent malic enzyme.
335 Moreover, convergent evolution in several key amino acid residues of PEPC has been suggested to be
336 associated with the origin of both $C_4$ and CAM (Christin et al. 2007; Christin et al. 2014; Goolsby et al.
337 2018). A recent study has also confirmed the occurrence of multiple rounds of duplication within the
338 major PEPC paralog (PEPC1E1) in the ancestral Portulacineae (Christin et al. 2014). The results
339 presented here provide additional evidence that gene family expansion may play an important role in
340 some aspects of photosynthetic pathway evolution.

341

342 **Lineage-specific gene expansions associated with adaptive traits**

343 In addition to investigating genes that experienced significant expansion, we also conducted Gene
344 Ontology (GO) overrepresentation analyses on genes duplicated at biologically significant nodes
345 (nodes 15, 17, 19, 26, 27, 35, 39, 50, and 71; fig. 1). Of genes duplicated at each clade, 30%–55% had
346 corresponding *Arabidopsis* locus IDs and almost all of these could be mapped in PANTHER (table S2).

347 We identified GO overrepresentation at three nodes (15, 27 and 35) with the Didiereoideae (node 27)
348 exhibiting the most. This may be the result of a more recent WGD. Overrepresented GOs reflected
349 diverse functional classes: genes belonging to "calcium ion binding" (GO:0005509) at the origins of
350 the Montiaceae (except *Phemeranthus parviflorus*) and the Didiereoideae, genes belonging to "sulfur
351 compound metabolic process" (GO:0006790) at the origin of Didiereoideae, and genes involved in
352 "Hydrolase activity, acting on ester bonds" (i.e., esterase, GO:0016788, table S3) at the origin of
353 Cactaceae. Of particular note are the genes belonging to the "sulfur compound metabolic process" (e.g.,
354 *AtAHL*, which belongs to the Arabidopsis HAL2-like gene family, Gil-Mascarell et al. 1999; and
355 *NIFS1*, which belongs to the cysteine desulfurase family, Schmidt 2005) as sulfur-bearing evaporite
356 compounds (principally gypsum) are commonly found in soils in areas with low rainfall and high
357 evaporation rates (Watson 1979) where many Portulacineae lineages have diversified. Previous studies
358 have also connected primary sulfur metabolism (e.g., sulfate transport in the vasculature, its
359 assimilation in leaves, and the recycling of sulfur-containing compounds) with drought stress responses
360 (e.g., Chan et al. 2013). Consequently, duplication and overrepresentation of these sulfur metabolic
361 genes could potentially be evidence of adaptation to corresponding harsh environments of hot and cold
362 deserts (e.g., Lee et al. 2011). While it is tempting to speculate that these duplications have been
363 maintained through adaptation under selective pressure, we are aware that such 'duplication' events
364 may simply be the remnants of ancient polyploidy or segmental duplication. In the absence of
365 functional characterization, it is necessary to interpret all GO analyses with caution.
366
367 **Targeted analyses of drought and cold associated genes**
368 We examined 29 functionally annotated genes identified to be involved in drought and cold tolerance
369 (table 2), as several lineages have repeatedly colonized cold and dry environments. Most of these genes
370 experienced significantly higher duplications than expected from WGDs including within
371 Didiereoideae, Montiaceae and Cactaceae (figs. S6 and S7). Several of these genes may be good targets
372 for future study. For example, WIN1 (SHN1) proteins, transcription factors associated with epicuticular
373 wax biosynthesis that increase leaf surface wax content, showed duplications in both Didiereoideae and
374 Montiaceae (table 2). Six homologs that are part of the calcium-dependent protein kinase (CDPK or
375 CPK) gene family are known to be involved in drought stress regulation (Geiger et al. 2010; Brandt et
376 al. 2012) and experienced several duplications: four at the origin of Didiereoideae, one within
377 Montiaceae, one at Portulacaceae, and two within the Cactaceae.
378 These 29 genes also exhibited variable lineage-specific positive selection within Portulacineae
379 (table 2). The strongest signals of positive selection were all found in genes associated with drought

380 and/or cold tolerance in the ABA signaling pathway (such as the NAC10/29 and WRKY33 TFs, the
381 CDPK18 and the PP2Cs, e.g., Golldack et al. 2014; Nakashima et al. 2014; Huang et al. 2015; Li et al.
382 2017), a central regulator of abiotic stress resistance. Interestingly, the ABI1 and PP2CA genes, which
383 encode two proteins of the PP2Cs family, were among the genes with the highest number of lineages
384 under positive selection. As negative regulators, PP2Cs can regulate many ABA responses, such as
385 stomatal closure, osmotic water permeability, drought-induced resistance, seed germination, and cold
386 acclimation (Gosti et al. 1999; Merlot et al. 2001; Mishra et al. 2006). In addition to PP2C genes, the
387 positive regulators SNF1-related protein kinase 2s (SnRK2s) are also related to ABA signaling
388 (Hubbard et al. 2010). While we did not detect many lineages under positive selection in the three
389 SnRK2s genes (i.e., SnRK2.4, 2.5, 2.6, table 2), we found that they experienced ancient duplications at
390 the origin of Portulacineae or earlier and recent duplications within Montiaceae, Didiereaceae, or
391 Cactaceae (nodes 17, 27, 39; table 2). Positive selection and gene family expansion in these gene
392 families within Cactaceae, Didieraceae, and Montiaceae suggests that these genes warrant further
393 investigation for their potential role in the evolution of adaptations to challenging environmental
394 conditions.
395 While several of these genes experienced duplications within clades of the Portulacinieae, many
396 occurred at the origin of Portulacineae or earlier (table 2, table S4 and fig. S7). This suggests that the
397 ability for the Caryophyllales, and the Portulacineae, to repeatedly adapt to harsh environments may
398 have arisen early in the evolution of the clade and that the predisposition to become adapted to these
399 environments may be the result of the early diversification of the gene families.
400
401 **Conclusions**
402 WGDs (i.e., paleopolyploidy) have played critical roles in major events of plant evolution, but the
403 nature and scale of their influence on macroevolutionary patterns and within individual clades is still
404 debated. Our results suggest that WGDs in the Portulacineae were associated with several evolutionary
405 patterns and processes. In addition to identifying significant gene-tree/species-tree discordance
406 immediately after extensive gene duplication events, suggesting an association between rapid
407 speciation and gene duplication and loss, we also found evidence for the association of WGDs and
408 shifts into different environments and climates (Smith et al. 2018a). While this alone does not suggest
409 specific cause, it suggests one means by which WGD can impact adaptation in plants and, perhaps,
410 why some clades exhibit diversification shifts after WGDs. However, we did not detect association
411 between WGD and diversification rate.

412 We also examined expanded gene families, GO over-representation, and selection within genes
413 to determine whether potential molecular evolutionary patterns corresponded to adaptation to extreme
414 environments. While these results provide essential resources for further and deeper examination with
415 more detailed methods, they also highlight the major limitations in resources for non-model organisms.
416 They rely on annotations based on model organisms that may be distantly related to the species of
417 interest. While the gene itself may exhibit interesting patterns, the associated GO label may or may not
418 carry much meaning and may have different functions in these different lineages. Despite these
419 limitations, however, these analyses identified gene families that should be explored further, regardless
420 of the accuracy of the GO label, as the patterns of selection and duplication alone suggest their
421 importance.
422 Finally, our findings, contribute to a growing literature suggesting the importance of gene tree
423 examination along with species tree construction (Smith SA et al. 2015; Brown and Thomson 2017;
424 Shen et al. 2017; Walker et al. 2018b). Gene tree conflict is common and needs to be incorporated into
425 our analyses of molecular and macro-evolution. The conflicts within these data likely hold interesting
426 evolutionary questions and answers that should be analyzed in more detail.
427
428 **Material and Methods**
429
430 **Taxon sampling, transcriptome generation and homology inference**
431 Sixty-eight ingroup species were included in this study, representing all families within the
432 Portulacineae (Anacampserotaceae, Cactaceae, Basellaceae, Didiereaceae, Montiaceae, Talinaceae, and
433 Portulacaceae) *sensu* APG IV (Angiosperm Phylogeny Group, 2016) except for Halophytaceae. We
434 also included 14 Caryophyllales species as outgroups, including seven taxa of Molluginaceae (tables
435 S5 and S6). Of the 82 transcriptomes included in the study, 47 were newly generated following the
436 protocol of Yang et al. (2017). We followed the pipeline of Yang and Smith (2014) with minor
437 modifications (available at https://bitbucket.org/ningwang83/Portulacineae) to infer homologs. In total,
438 we obtained 8,592 final homolog clusters, which are used in orthology inference, gene duplication
439 analyses and GO annotation. More details about transcriptome generation and homology inference can
440 be found in Supplementary Material and Methods.
441
442 **Orthology Inference and Species Tree Estimation**
443 We used the rooted tree (RT) method in Yang and Smith (2014) to extract orthologs from homolog
444 trees with *Beta vulgaris* as the outgroup. Orthologous clades with more than 30 ingroup taxa were

445 retained (fig. S8). We extracted orthologs with > 75 taxa, aligned and cleaned them with MAFFT v7
446 (genafpair –maxiterate 1000, Katoh and Standley 2013) and Phyutility (-clean 0.3, Smith and Dunn
447 2008), and inferred a ML tree for each using RAxML v8.1.22 (Stamatakis 2015) with GTR+Γ model.
448 After removing taxa with a terminal branch length longer than 0.1 and 10 times greater than sister clade
449 from 58 orthologs, 841 orthologs including at least 77 taxa and 500 aligned DNA characters were re-
450 aligned with PRANK v.140110 (Löytynoja and Goldman 2008) using default settings and trimmed in
451 Phyutility (-clean 0.3).
452 The species tree was inferred using two methods. First, a concatenated matrix was built with the
453 841 orthologs and a ML tree was estimated by using RAxML with the GTR+Γ model partitioned by
454 gene. Node support was evaluated by 200 fast bootstrap replicates. Second, an MQSST tree was
455 estimated in ASTRAL 4.10.12 (Mirarab et al. 2014) using ML gene trees, with uncertainty evaluated
456 by 200 bootstrap replicates using a two-stage multilocus bootstrap strategy (Seo 2008).
457
458 **Assessing conflicts among gene trees**
459 We assessed conflict by mapping the 841 rooted ortholog trees onto the species tree topology with
460 Phyparts (Smith SA et al. 2015). Trees were rooted on *Beta vulgaris* if present, using *Limeum*
461 *aethiopicum* and/or *Stegnosperma halimifolium* if *Beta* was not present, and with *Sesuvium*
462 *portulacastrum*, *Delosperma echinatum*, *Anisomeria littoralis* and *Guapira obtusata,* if the first three
463 species were not present. We only considered nodes with >= 70% bootstrap support in the gene trees.
464 We conducted likelihood comparisons in more detail as in Walker et al. (2018b) for three areas of the
465 tree: 1) Cactaceae and relatives, 2) the early diverging branch within Cactaceae (i.e., *Leuenbergeria*
466 clade), and 3) with the placement of Basellaceae and Didiereaceae. We calculated likelihood scores for
467 each gene tree constrained to alternative resolutions and compared scores. We considered a higher ln $L$
468 for one topology as a sign of support, and $|\Delta \ln L|$ (the difference in ln $L$ value) > 2 as a sign of
469 significant support (as in Walker et al. 2018b). Potential outlying genes that exhibited extreme
470 deviations in likelihood were examined for lineage specific positive selection. We analyzed potential
471 hybridization with PhyloNet (Than et al. 2008) and $D_{FOIL}$ (Pease and Hahn 2015). PhyloNet does not
472 run well on large datasets so we pruned the gene trees to include 5 taxa from Cactaceae, one from each
473 ingroup family, with the *Limeum aethiopicum* as an outgroup. Species were selected based on their
474 gene occupancy. We conducted Maximum Pseudo Likelihood analyses with 10 maximum reticulations.
475 $D_{FOIL}$ analyses can only analyze 5-taxon trees, so we reduced gene trees to four species with the highest
476 gene occupancy for the clade of interest and *Pharnaceum exiguum* as an outgroup. For example, we

477 choose two species from Cactaceae and one from each of Portulacaceae and Anacampserotaceae, to
478 test the controversial relationships among these families.
479
480 **Inference of gene and genome duplication**
481 To infer gene duplication, we considered nodes in gene trees with SH-like support > 80. We extracted
482 8,332 rooted clusters with >= 30 taxa, and conducted Phyparts analyses. We also conducted Ks plot
483 analyses following the same process of Yang et al. (2015). Briefly, we reduced highly similar PEP
484 sequences (by CD-HIT: -c 0.99 -n 5, Fu et al. 2012) for each of the 82 species, and conducted an all-
485 by-all BLASTP (-evalue = 10, -max_target_seq = 20) and removed highly divergent hits with < 20%
486 similarity (pident) or < 50 aligned amino acids (nident). Sequences with ten or more hits were also
487 removed to eliminate large gene families. We then used the PEP and corresponding CDS from
488 paralogous pairs to calculate Ks values using the pipeline
489 https://github.com/tanghaibao/biopipeline/tree/master/synonymous_calculation. Peaks were identified
490 by eye as in Yang et al. (2015).
491 To determine whether a potential WGD event occurred before or after a speciation event, we
492 calculated between-species Ks distribution using orthologous gene pairs. The procedure is similar as
493 above, except that a reciprocal BLASTP was carried out between two species instead of an all-by-all
494 BLASTP within one taxon. Between-species Ks plots were compared to within-species Ks plots to
495 determine the relative timing of the WGD and speciation event.
496
497 **Reconstruction of climate occupancy and diversification rate**
498 To test whether WGD were associated with climate niche shift and/or diversification rate shift, we
499 reconstructed ancestral climate occupancy and diversification on a species level phylogeny of
500 Portulacineae. First, sequence data from NCBI were gathered using PyPHLAWD (Smith and Brown
501 2018). To increase the accuracy of phylogenetic inferences, major clades were constrained based on the
502 transcriptomic results here and were conducted on Molluginaceae and Cactineae (i.e., Portulacineae)
503 individually and combined. ML and 100 bootstrap trees were constructed with RAxML. For
504 divergence time estimation, we selected five genes from the transcriptomic dataset using SortaDate
505 (Smith et al. 2018b). A time tree, hereafter 'small tree', was then built with them in BEAST 1.8.3 using
506 13 secondary time calibrations (table S7) from Arakaki et al. (2011). The species-level tree, hereafter
507 'big tree', as well as the 93 bootstrap trees that were consistent among family level relationships in the
508 'small tree' were then dated in treePL (Smith and O'Meara 2012) using the dates from the 'small tree'
509 as constraints.

510 Reconstruction of ancestral climate occupancy state and diversification rate followed the
511 methods of Smith et al. (2018a). Briefly, we extracted the climate occupancy data (COD) for the
512 Portulacineae and Molluginaceae species from GBIF based on the taxa from the 'big tree'. Mean
513 climate values of Bioclim 1 (mean annual temperature), 12 (mean annual precipitation), and principal
514 component axis 1 (based on a PCA of the set of full bioclimatic variables) were calculated for each
515 taxon. Both ancestral states and Brownian motion rates of evolution were compared between WGD
516 nodes and their sister clades. We reconstructed diversification using MEDUSA (Alfaro et al., 2009;
517 Pennell et al., 2014) on the 'small tree' with the estimated number of species for each taxa group and
518 the 'big tree'.
519
520 **Annotation of expanded gene families**
521 To investigate gene function and GO (gene ontology) term overrepresentation, we conducted several
522 analyses. We conducted BLASTX analyses against the nonredundant protein NCBI database by using a
523 PEP sequence from each of the top 20 genes with the highest total number of tips. We also identified
524 clades with large number of gene duplications [i.e., > 160 (2%) gene duplications], and conducted
525 BLASTN analyses (-evalue 10) of these against the database of *A. thaliana* (release-34 EnsemblPlants).
526 GO terms and IDs were recorded. We conducted GO overrepresentation test (i.e., overrepresentation of
527 genes in a specific functional category, http://www.pantherdb.org) using genes duplicated at the origin
528 of Portulacineae (Node 13) as background to detect potential overrepresentation of genes that
529 duplicated more recently. We corrected for multiple tests using a Bonferroni correction.
530
531 **Identification of lineage-specific selection on stress response genes**
532 We carried out selection analyses on a targeted set of 29 genes that were present in our data and are
533 known to be associated with cold and/or drought adaptation. We repeated the homolog filtering as
534 indicated above three times to reduce potential assembly or clustering errors. CDS sequences were
535 aligned based on the corresponding peptide alignments using phyx (Brown et al. 2017). We further
536 cleaned CDS alignments by combining sequences with identical overlapping segments to eliminate the
537 potential assemble redundancy, which, although uncommon, may cause error in these analyses. After
538 trimming the CDS alignment (Phyutility -clean 0.1), a ML tree was built for each homolog in RAxML
539 with GTR+Γ model and rooted with outgroup species. Lineage specific selection was calculated for
540 each homologous gene in HyPhy v2.2.4 (Pond and Muse 2005) using an adaptive branch-site random
541 effects likelihood test (aBSREL, Smith MD et al. 2015), which automatically infers an appropriate
542 model among branches. The optimal number of ω categories was selected according to AICc scores.

543 The branches with episodic positive selection that show significant proportion of sites with $\omega > 1$ were
544 chosen with $p < 0.05$ after applying the Holm-Bonferroni multiple testing correction.
545 To test whether the 29 targeted genes associated with cold/drought adaptation experienced
546 significant more duplications at certain nodes, we randomly selected 29 genes from the final
547 homologous clusters (with an additional overlap checking to be compatible with the treatment of the
548 targeted 29 genes) and calculate the number of genes duplicated. We repeated this process for 1000
549 times and plotted the frequency distribution of the number of duplicated genes for each node (fig. S7).
550

551 **Supplementary Material**

552 Supplementary data are available at Molecular Biology and Evolution online.
553

554 **Acknowledgments**

555 The authors thank Wynn Anderson, the Bureau of Land Management, the US Forest Service, and the
556 staff of Desert Botanical Garden, Sukkulenten-Sammlung Zürich, Cambridge University Botanic
557 Garden, Missouri Botanical Garden, and the Oberlin College Greenhouse for permission to collect
558 specimens, and thank Hilda Flores, Helga Ochoterena, and Norman Douglas for help with collecting.
559 The authors also thank Jeet Sukumaran for constructive discussion on gene tree discordance. Special
560 thanks to Oscar Vargas for a thorough review of the paper. This work was supported by NSF DEB
561 1354048 to SAS and NSF DEB 1352907 to MJM.
562

563 **References**

564 Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of
565 Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic
566 confidence. *Mol Biol Evol.* 20(2):255–266.
567 Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ. 2009. Nine
568 exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad*
569 *Sci USA.* 106(32):13410–13414.
570 Anderson EF. 2001. The Cactus Family, Pentland, Oregon: Timber Press.
571 Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the
572 orders and families of flowering plants: APG IV. *Bot J Linean Soc.* 181(1):1–20.
573 Anton AM, Hernández-Hernández T, De-Nova A, Sosa V. 2014. Evaluating the phylogenetic position of the
574 monotypic family Halophytaceae (Portulacinae, Caryophyllales) based on plastid and nuclear molecular
575 data sets. *Bot Sci.* 92(3):351–361.

576 Applequist WL, Wallace RS. 2000. Phylogeny of the Madagascan endemic family Didiereaceae. *Plant Syst Evol.*
577 221(3):157–166.
578 Applequist WL, Wallace RS. 2003. Expanded circumscription of Didiereaceae and its division into three
579 subfamilies. *Adansonia* 25(1):13–16.
580 Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011.
581 Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad*
582 *Sci USA.* 108(20):8379–8384.
583 Bárcenas RT, Yesson C, Hawkins JA. 2011 Molecular systematics of the Cactaceae. *Cladistics* 27:470–489
584 Barthlott W, Hunt DR. 1993. Cactaceae. In Flowering Plants Dicotyledons Springer Berlin Heidelberg. p. 161–
585 197.
586 Borland AM, Griffiths H, Hartwell J, Smith JA. 2009. Exploiting the potential of plants with crassulacean acid
587 metabolism for bioenergy production on marginal lands. *J Exp Bot.* 60(10):2879–2896.
588 Brandt B, Brodsky DE, Xue S, Negi J, Iba K, Kangasjärvi J, Ghassemian M, Stephan AB, Hu H, Schroeder JI.
589 2012. Reconstitution of abscisic acid activation of SLAC1 anion channel by CPK6 and OST1 kinases
590 and branched ABI1 PP2C phosphatase action. *Proc Natl Acad Sci USA.* 109(26):10593–10598.
591 Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R. 2004. Polyploidy in arctic
592 plants. *Bio J Linnean Soc.* 82:521–536.
593 Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GK, Moore MJ, Smith
594 SA. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in
595 Caryophyllales. *New Phytol.* 207:1170–1180.
596 Brown JM, Thomson RC. 2017. Bayes factors unmask highly variable information content, bias, and extreme
597 influence in phylogenomic analyses. *Syst Biol.* 66(4):517–530.
598 Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33(12):1886–1888.
599 Bruyns PV, Oliveira-Neto M, Melo-de-Pinna GF, Klak C. 2014. Phylogenetic relationships in the Didiereaceae
600 with special reference to subfamily Portulacarioideae. *Taxon* 63(5):1053–1064.
601 Butterworth CA, Cota-Sanchez JH, Wallace RS. 2002. Molecular systematics of tribe Cacteae (Cactaceae:
602 Cactoideae): a phylogeny based on rpl16 intron sequence variation. *Syst Bot.* 27(2):257–270.
603 Butterworth CA, Wallace RS. 2005. Molecular phylogenetics of the leafy cactus genus *Pereskia* (Cactaceae).
604 *Syst Bot.* 30(4):800–808.
605 Chan KX, Wirtz M, Phua SY, Estavillo GM, Pogson BJ. 2013. Balancing metabolites in drought: the sulfur
606 assimilation conundrum. *Trends Plant Sci.* 18(1):18–29.
607 Christin PA, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK,
608 Hancock L, Edwards EJ. 2014. Shared origins of a key enzyme during the evolution of C4 and CAM
609 metabolism. *J Exp Bot.* 65:3609–3621.
610 Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 Photosynthesis Evolved in Grasses via
611 Parallel Adaptive Genetic Changes. *Curr Biol.* 17:1241–1247.

612 Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K,
613 Barakat A, Albert VA. 2006. Widespread genome duplications throughout the history of flowering
614 plants. *Genome Res.* 16(6):738–749.
615 Cushman JC. 2001. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments.
616 *Plant Physiol.* 127:1439–1448.
617 Edwards EJ, Nyffeler R, Donoghue MJ. 2005. Basal cactus phylogeny: implications of *Pereskia* (Cactaceae)
618 paraphyly for the transition to the cactus life form. *Am J Bot.* 92(7):1177–1188.
619 Edwards EJ, Ogburn RM. 2012. Angiosperm responses to a low-CO2 world: CAM and C4 photosynthesis as
620 parallel evolutionary trajectories. *Int J of Plant Sci.* 173(6):724–733.
621 Edwards G, Walker D. 1983. C3, C4: mechanisms, and cellular and environmental regulation, of photosynthesis.
622 University of California Press.
623 Emms DM, Covshoff S, Hibberd JM, Kelly S. 2016. Independent and parallel evolution of new genes by gene
624 duplication in two origins of C4 photosynthesis provides new insight into the mechanism of phloem
625 loading in C4 species. *Mol Biol Evol.* 33(7):1796–1806.
626 Ferreyra MLF, Andreo CS, Podestá FE. 2003. Purification and physical and kinetic characterization of a
627 photosynthetic NADP-dependent malic enzyme from the CAM plant Aptenia cordifolia. *Plant Sci.*
628 164(1):95–102.
629 Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data.
630 *Bioinformatics* 28(23):3150–3152.
631 Geiger D, Scherzer S, Mumm P, Marten I, Ache P, Matschi S, Liese A, Wellmann C, Al-Rasheid KA, Grill E,
632 Romeis T, Hedrich R. 2010. Guard cell anion channel SLAC1 is regulated by CDPK protein kinases
633 with distinct Ca2+ affinities. *Proc Natl Acad Sci USA.* 107(17):8023–8028.
634 Gibson AC, Nobel PS. 1990. The cactus primer. Harvard University Press
635 Gibson AC. 1996. Structure–function relations of warm desert plants. New York, NY, USA: Springer.
636 Gil-Mascarell R, López-Coronado JM, Bellés JM, Serrano R, Rodríguez PL. 1999. The *Arabidopsis* HAL2 like
637 gene family includes a novel sodium‐sensitive phosphatase. *Plant J.* 17(4):373–383.
638 Golldack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: unraveling the
639 signaling networks. *Front Plant Sci.* 5:151.
640 Goolsby EW, Moore AJ, Hancock LP, De Vos JM, Edwards EJ. 2018. Molecular evolution of key metabolic
641 genes during transitions to C4 and CAM photosynthesis. *Am J Bot.* 105(3): 602–613.
642 Gosti F, Beaudoin N, Serizet C, Webb AA, Vartanian N, Giraudat J. 1999. ABI1 protein phosphatase 2C is a
643 negative regulator of abscisic acid signaling. *Plant Cell* 11(10):1897–1909.
644 Griffith MP, Porter JM. 2009. Phylogeny of Opuntioideae (Cactaceae). *International J Plant Sci.* 170:107–116.

645 Hernández-Hernández T, Hernández HM, De-Nova JA, Puente R, Eguiarte LE, Magallón S. 2011. Phylogenetic
646 relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot.*
647 98(1):44–61.
648 Hernández-Ledesma P, Berendsohn WG, Borsch T, Mering SV, Akhani H, Arias S, Castañeda-Noa I, Eggli U,
649 Eriksson R, Flores-Olvera H, Fuentes-Bazán S. 2015. A taxonomic backbone for the global synthesis of
650 species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45(3):281–383.
651 Hernández-Hernández T, Brown JW, Schlumpberger BO, Eguiarte LE, Magallón S. 2014. Beyond aridification:
652 multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New*
653 *Phytol.* 202(4):1382–1397.
654 Hochstrasser M. 2009. Origin and function of ubiquitin-like proteins. *Nature* 458(7237):422–429.
655 Huang Q, Wang Y, Li B, Chang J, Chen M, Li K, Yang G, He G. 2015. TaNAC29, a NAC transcription factor
656 from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC Plant Biol.* 15(1):268.
657 Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI. 2010. Early abscisic acid signal transduction
658 mechanisms: newly discovered components and newly emerging questions. *Genes Dev.* 24(16):1695–
659 1708.
660 Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H,
661 Soltis PS, Soltis DE. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–
662 100.
663 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
664 performance and usability. *Mol Biol Evol.* 30(4):772–780.
665 Kiang JG, Tsokos GC. 1998. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology.
666 *Pharmacol Ther.* 80(2):183–201.
667 Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in
668 response to salt, osmotic, and cold stress. *Plant Physiol.* 130(4):2129–2141.
669 Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome
670 duplication events on diversification rates in angiosperms. American journal of botany.
671 Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson
672 DW, McCombie WR, Martienssen RA. 2011. A functional phylogenomic view of the seed plants. *PLoS*
673 *Genet.* 7(12):e1002411.
674 Li H, Chang J, Zheng J, Dong Y, Liu Q, Yang X, Wei C, Zhang Y, Ma J, Zhang X. 2017. Local melatonin
675 application induces cold tolerance in distant organs of *Citrullus lanatus* L. via long distance
676 transport. *Sci Rep.* 7:40858.
677 Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA
678 sequence data. *Evolution* 62(8):2080–2091.

679 Liu L, Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a
680 species phylogeny using estimated gene tree distributions. *Syst Biol.* 56(3):504–514.
681 Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and
682 evolutionary analysis. *Science* 320(5883):1632–1635.
683 Mauseth JD. 2006. Structure–function relationships in highly modified shoots of Cactaceae. *Ann Bot.* 98(5):901–
684 926.
685 Merlot S, Gosti F, Guerrier D, Vavasseur A, Giraudat J. 2001. The ABI1 and ABI2 protein phosphatases 2C act
686 in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* 25(3):295–303.
687 Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale
688 coalescent-based species tree estimation. *Bioinformatics* 30(17):i541–i548.
689 Mishra G, Zhang W, Deng F, Zhao J, Wang X. 2006. A bifurcating pathway directs abscisic acid effects on
690 stomatal closure and opening in Arabidopsis. *Science* 312(5771):264–266.
691 Moore A, de Vos JM, Hancock LP, Goolsby E, Edwards EJ. 2017. Targeted Enrichment of Large Gene Families
692 for Phylogenetic Inference: Phylogeny and Molecular Evolution of Photosynthesis Genes in the
693 Portullugo (Caryophyllales). *Syst Biol.* 67(3):367–383..
694 Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. NAC transcription factors in
695 plant abiotic stress responses. *Biochim Biophys Acta* 1819(2):97–103.
696 Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 2014. The transcriptional regulatory network in the
697 drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front*
698 *Plant Sci.* 5:170.
699 Nyffeler R, Eggli U. 2010a. Disintegrating Portulacaceae: a new familial classification of the suborder
700 Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* 59(1):227–240.
701 Nyffeler R, Eggli U. 2010b. A farewell to dated ideas and concepts –molecular phylogenetics and a revised
702 suprageneric classification of the family Cactaceae. *Schumannia* 6:109–149.
703 Nyffeler R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from
704 trnK/matK and trnL-trnF sequences. *Am J Bot.* 89:312–326.
705 Nyffeler R. 2007. The closest relatives of cacti: insights from phylogenetic analyses of chloroplast and
706 mitochondrial sequences with special emphasis on relationships in the tribe Anacampseroteae. *Am J Bot.*
707 94(1):89–101.
708 Ocampo G, Columbus JT. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales), including
709 insights into photosynthetic diversification and historical biogeography. *Am J Bot.* 97(11):1827–1847.
710 Ogburn RM, Edwards EJ. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm
711 clade Montiaceae. *Mol Phylogenet Evol.* 92:181–192.
712 Parks MB, Nakov T, Ruck EC, Wickett NJ, Alverson AJ. 2018. Phylogenomics reveals an extensive history of
713 genome duplication in diatoms (Bacillariophyta). *Am J Bot.* 105(3):330–347.

714 Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Sys Bio.* 64(4):
715 651–662.
716 Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification.
717 *Hydrol Earth Syst Sci.* 11:1633–1644.
718 Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014. geiger
719 v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees.
720 Bioinformatics, 30(15):2216–2218.
721 Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol*
722 *Evol.* 21:1455–1458.
723 Pond, S.L.K. and Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. In: Statistical methods in
724 molecular evolution. Springer, New York, NY. p. 125–181.
725 Qin F, Sakuma Y, Li J, Liu Q, Li YQ, Shinozaki K, Yamaguchi-Shinozaki K. 2004. Cloning and functional
726 analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea*
727 *mays* L. *Plant Cell Physiol.* 45(8):1042–1052.
728 Schmidt A. 2005. Metabolic background of H2S release from plants, in: L.J. De Kok, E.S. Chnug (Eds.),
729 Proceedings of the 1st Sino-German Workshop on Aspects of Sulfur Nutrition of Plants, 17th End,
730 Braunschweig, Germany, pp. 121–129.
731 Seo TK. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol Biol Evol.*
732 25(5):960–971.
733 Shen X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a
734 handful of genes. *Nature Ecol Evol.* 1:1–10.
735 Silvera K, Winter K, Rodriguez BL, Albion RL, Cushman JC. 2014. Multiple isoforms of phosphoenol-pyruvate
736 carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid
737 metabolism. *J Exp Bot.* 65(13):3623–3636.
738 Smith JAC, Winter K. 1996. Taxonomic distribution of crassulacean acid metabolism. In: Winter K, Smith JAC,
739 eds. Crassulacean acid metabolism: biochemistry, ecophysiology and evolution. Berlin/Heidelberg:
740 Springer-Verlag. p. 427–436.
741 Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an
742 adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol*
743 *Biol Evol.* 32(5):1342–1353.
744 Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. Am J Bot. 105(3): 302–314.
745 Smith SA, Brown JW, Walker JF. 2018b. So many genes, so little time: A practical approach to divergence-time
746 estimation in the genomic era. PLoS ONE 13(5): e0197433.
747 Smith SA, Brown JW, Yang Y, Bruenn R, Drummond CP, Brockington SF, Walker JF, Last N, Douglas NA,
748 Moore MJ. 2018a. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.* 217(2):836–
749 854.

750 Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data.
751 *Bioinformatics* 24(5):715–716.
752 Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict,
753 concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* 15(1):150.
754 Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large
755 phylogenies. *Bioinformatics* 28:2689–2690.
756 Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB,
757 Moore MJ, Carlsward BS, Bell CD. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot.*
758 98(4):704–730.
759 Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588.
760 Soltis PS, Soltis DS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl*
761 *Acad Sci USA*. 97:7051–7057.
762 Stamatakis A. 2015. Using RAxML to infer phylogenies. *Curr Protoc Bioinformatics*. 51:6–14.
763 Stebbins GL. 1971. Chromosomal evolution in higher plants. Addison Wesley, London. p. 216.
764 Stoughton TR, Jolles DD, O'Quinn RL. 2017. The western spring beauties, *Claytonia lanceolata* (Montiaceae): a
765 review and revised taxonomy for California. *Syst Bot.* 42(2):283–300.
766 Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate
767 evolutionary relationships. *BMC bioinformatics* 9(1):322.
768 van den Bergh E, Külahoglu C, Bräutigam A, Hibberd JM, Weber AP, Zhu XG, Schranz ME. 2014. Gene and
769 genome duplications and the origin of C4 photosynthesis: Birth of a trait in the Cleomaceae. *Curr Plant*
770 *Biol.* 1:2–9.
771 Vandeleur RK, Mayo G, Shelden MC, Gilliham M, Kaiser BN, Tyerman SD. 2009. The role of plasma
772 membrane intrinsic protein aquaporins in water transport through roots: diurnal and drought stress
773 responses reveal different strategies between isohydric and anisohydric cultivars of grapevine. *Plant*
774 *Physiol.* 149(1):445–460.
775 Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions revisited.
776 *Mol Biol Evol.* 30(1):177–90.
777 Vázquez-Sánchez M, Terrazas T, Arias S, Ochoterena H. 2013. Molecular phylogeny, origin and taxonomic
778 implications of the tribe Cacteae (Cactaceae). *Sys Biodivers.* 11(1):103–116.
779 Walker JF, Yang Y, Feng T, Timoneda A, Mikenas J, Hutchinson V, Edwards C, Wang N, Ahluwalia S, Olivieri
780 J, et al. 2018a. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical
781 homology inference provide further insight to the evolution of Caryophyllales. *Am J Bot.* special issue
782 on the Plant Tree of Life, 105(3):446–462.
783 Walker JF, Brown JW, Smith, SA. 2018b. Analyzing contentious relationships and outlier genes in
784 phylogenomics. *Syst Biol.* syy043, https://doi.org/10.1093/sysbio/syy043.

785 Walker JF, Yang Y, Moore MJ, Mikenas J, Timoneda A, Brockington SF, Smith SA. 2017. Widespread
786 paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales.
787 *Am J Bot.* 104(6):858–867.
788 Wallace RS, Dickie SL. 2002. Systematic implications of chloroplast DNA sequence variation in subfam.
789 Opuntioideae (Cactaceae). *Succ. Pl. Res.* 6:9–24.
790 Watson A. 1979. Gypsum crusts in deserts. *J Arid Environ*. 2(1):3–21.
791 Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *Am J Bot.* 102:1753–1756.
792 Williams BP, Aubry S, Hibberd JM. 2012. Molecular evolution of genes recruited into C4
793 photosynthesis. *Trends Plant Sci.* 17(4):213–220.
794 Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of
795 polyploid speciation in vascular plants. *Proc Natl Acad Sci USA*. 106:13875–13879.
796 Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, Smith SA. 2018. Improved transcriptome
797 sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two
798 allopolyploidy events. *New Phytol.* 217(2):855–870
799 Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y,
800 Sage RF, Covshoff S, Hibberd JM, Nelson MN, Smith SA. 2015. Dissecting molecular evolution in the
801 highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol.* 32(8):2001–
802 2014.
803 Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage
804 genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol Biol Evol.* 31(11):3081–
805 3092.
806

807
808 Table 1. Portulacineae gene families with the highest total number of tips
809

| Cluster ID | Ntip | Ntaxa | Annotation |
|---|---|---|---|
| cluster1_2rr_1rr_2rr | 949 | 82 | tubulin beta chain [*Spinacia oleracea*] |
| cluster2_1rr_1rr_1rr | 871 | 82 | heat shock cognate 70 kDa protein-like [*Nicotiana tabacum*] |
| cluster4_1rr_1rr_1rr | 698 | 82 | Actin-11, NBD sugar-kinase HSP70 actin [*Chenopodium quinoa*] |
| cluster3_3rr_1rr_1rr | 670 | 81 | Polyubiquitin-A [*Triticum urartu*] |
| cluster5_1rr_1rr_1rr | 655 | 82 | plasma membrane H+ ATPase 9 [*Sesuvium portulacastrum*] |
| cluster8_1rr_1rr_1rr | 582 | 82 | tubulin alpha-3 chain [*Vitis vinifera*] |
| cluster12_1rr_1rr_1rr | 565 | 82 | phosphoenolpyruvate carboxylase [*Suaeda aralocaspica*] |
| cluster43_1rr_1rr_1rr | 488 | 81 | ubiquitin-conjugating enzyme E2 28 [*Solanum lycopersicum*] |
| cluster18_1rr_1rr_1rr | 472 | 82 | serine/threonine-protein kinase tricorner isoform X1 [*Beta vulgaris vulgaris*] |
| cluster22_1rr_1rr_1rr | 467 | 82 | NADP-dependent malic enzyme [*Hylocereus undatus*] |
| cluster15_1rr_1rr_1rr | 463 | 82 | transmembrane 9 superfamily member 3 [*Chenopodium quinoa*] |
| cluster17_1rr_1rr_1rr | 462 | 82 | auxin transporter-like protein 2 [*Chenopodium quinoa*] |
| cluster19_1rr_1rr_1rr | 461 | 82 | plasma membrane intrinsic protein 2 [*Sesuvium portulacastrum*] |
| cluster7_1rr_1rr_1rr | 456 | 81 | uridine-cytidine kinase C isoform X1 [*Beta vulgaris vulgaris*] |
| cluster26_1rr_1rr_1rr | 435 | 82 | S-adenosylmethionine synthase 1 [*Beta vulgaris vulgaris*] |
| cluster14_1rr_1rr_1rr | 467 | 82 | probable xyloglucan endotransglucosylase/hydrolase protein 23 [*Beta vulgaris vulgaris*] |
| cluster6_2rr_1rr_1rr | 530 | 81 | homeobox-leucine zipper protein ATHB-15 isoform X1 [*Beta vulgaris vulgaris*] |
| cluster9_1rr_1rr_1rr | 527 | 81 | expansin-A10-like [*Chenopodium quinoa*] |
| cluster16_1rr_1rr_1rr | 438 | 82 | ubiquitin carboxyl-terminal hydrolase 12 [*Beta vulgaris vulgaris*] |
| cluster37_1rr_1rr_1rr | 435 | 82 | 14-3-3 protein 10-like [*Chenopodium quinoa*] |

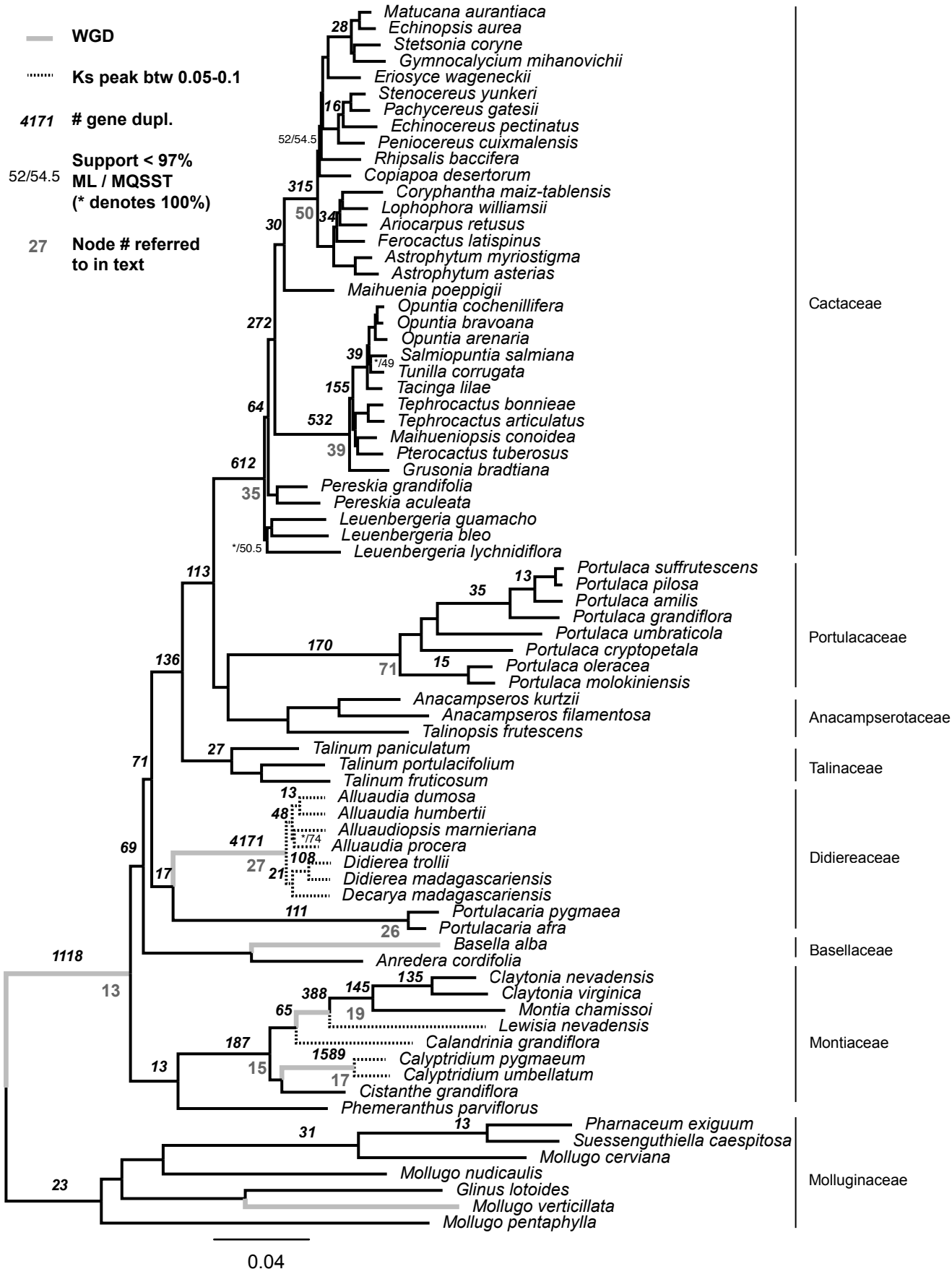
810 NOTE.—Ntip: number of tips, Ntaxa: number of taxa

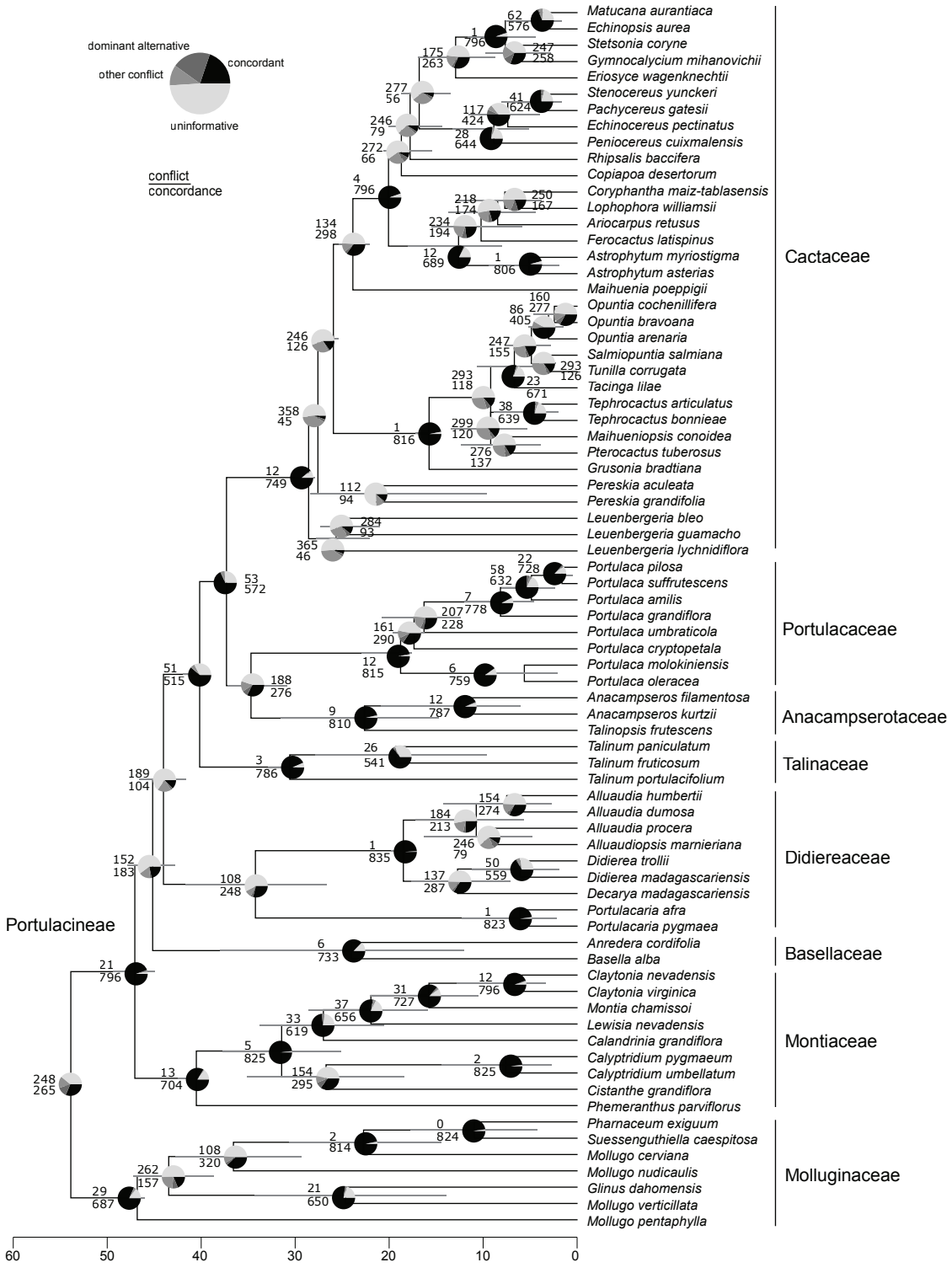
811 Table 2. Selection analyses on 29 genes associated with cold and drought responses.
812

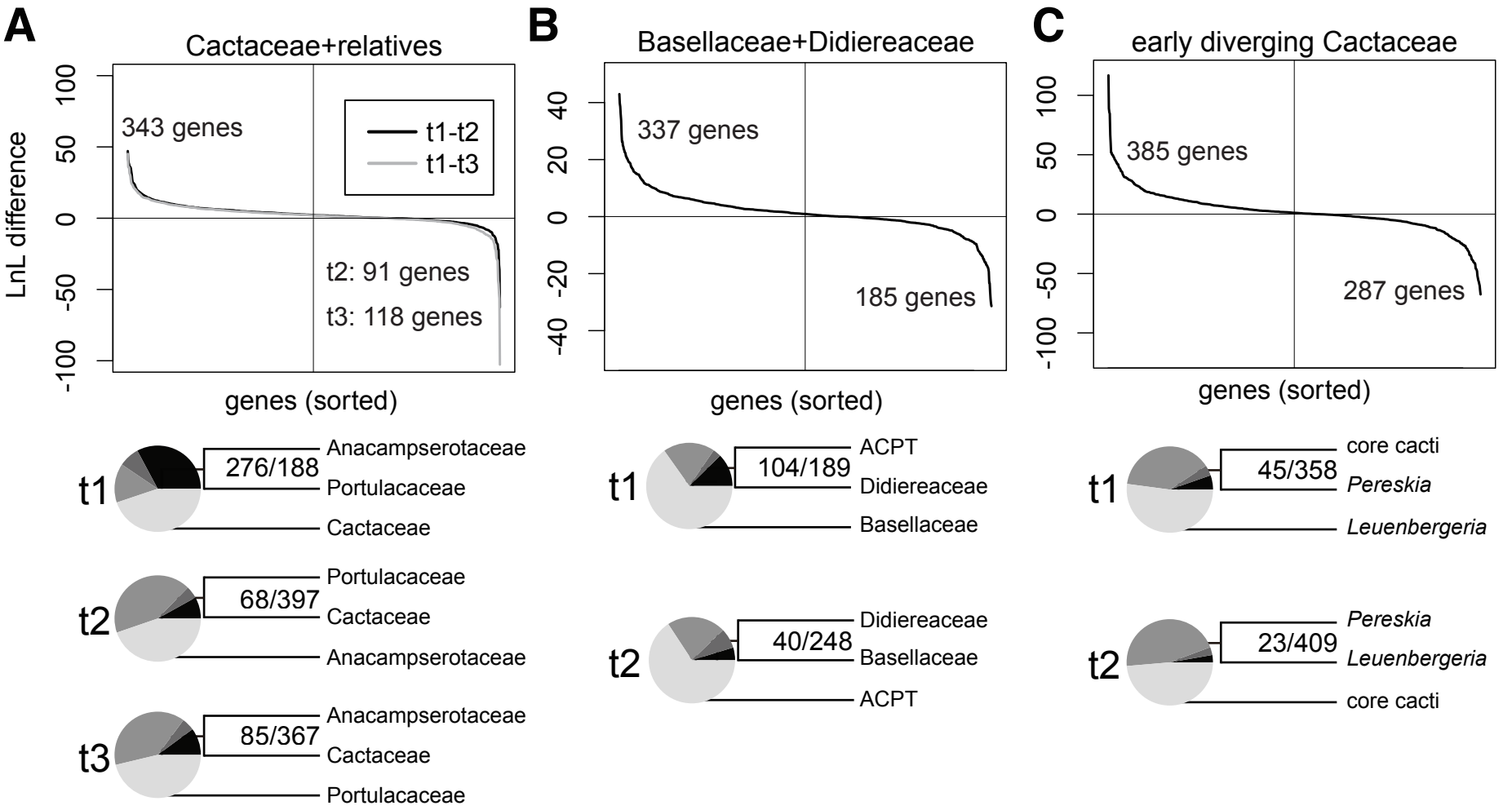
| Gene | TN | TS | Cac35 | Did27 | Mont15 | Port71 | SH80 (duplication) |
|---|---|---|---|---|---|---|---|
| **ABI1/PP2Cfamily** | 151 | 17 | 6 | 2 | 4 | 0 | 27 |
| NAC10 | 158 | 16 | 6 | 4 | 3 | 1 | 27,39 |
| WRKY33 | 165 | 13 | 4 | 2 | 3 | 1 | 17,27 |
| NAC29 | 163 | 13 | 7 | 2 | 2 | 0 | 13 |
| CDPK18 | 196 | 13 | 6 | 1 | 0 | 0 | 2,13,21,35,57 |
| **PP2C37 (PP2CA)** | 179 | 11 | 6 | 1 | 2 | 0 | 13,17,19,35 |
| HSP70 | 270 | 10 | 3 | 1 | 1 | 2 | 0,19,27,33,49,51,52 |
| ICE1 | 194 | 10 | 4 | 1 | 3 | 0 | 17,20,27,38 |
| HSP70 | 221 | 10 | 2 | 0 | 2 | 0 | 0,27,35 |
| ERF5 | 86 | 9 | 1 | 2 | 3 | 0 | 17,27,30 |
| HSP90-5 | 213 | 8 | 2 | 2 | 1 | 1 | 26,27,34 |
| CDPK1 | 85 | 7 | 4 | 2 | 1 | 0 | --- |
| SNRK2.5 | 156 | 5 | 2 | 1 | 0 | 1 | 39 |
| MYB124 (FLP) | 178 | 5 | 2 | 1 | 0 | 1 | 39 |
| HSP83 | 305 | 5 | 1 | 1 | 0 | 0 | 0,13,17,27,34,76 |
| CRPK1 | 177 | 5 | 3 | 2 | 0 | 0 | 13,20,27 |
| AREB1 | 92 | 4 | 1 | 1 | 1 | 0 | 17,27 |
| NAC2 | 157 | 4 | 2 | 0 | 0 | 1 | 27,39 |
| MYB60 | 87 | 4 | 2 | 0 | 0 | 1 | 27 |
| WIN1 (SHN1) | 79 | 4 | 0 | 0 | 0 | 1 | 17,27 |
| CDPK13 | 156 | 3 | 0 | 0 | 0 | 2 | 27 |
| CDPK26 | 117 | 3 | 1 | 1 | 0 | 1 | 0,17,27 |
| PLC2 | 108 | 3 | 1 | 0 | 1 | 0 | 17,19,35 |
| CDPK8 | 184 | 2 | 1 | 0 | 0 | 0 | 13,27,71 |
| CRLK1 | 170 | 2 | 2 | 0 | 0 | 0 | 27 |
| SNRK2.6 | 180 | 2 | 2 | 0 | 0 | 0 | 27,39 |
| SNRK2.4 | 125 | 1 | 0 | 0 | 0 | 0 | 17,27 |
| CDPK10 | 85 | 1 | 1 | 0 | 0 | 0 | 27 |
| ERF2 | 84 | 0 | 0 | 0 | 0 | 0 | 17,27 |

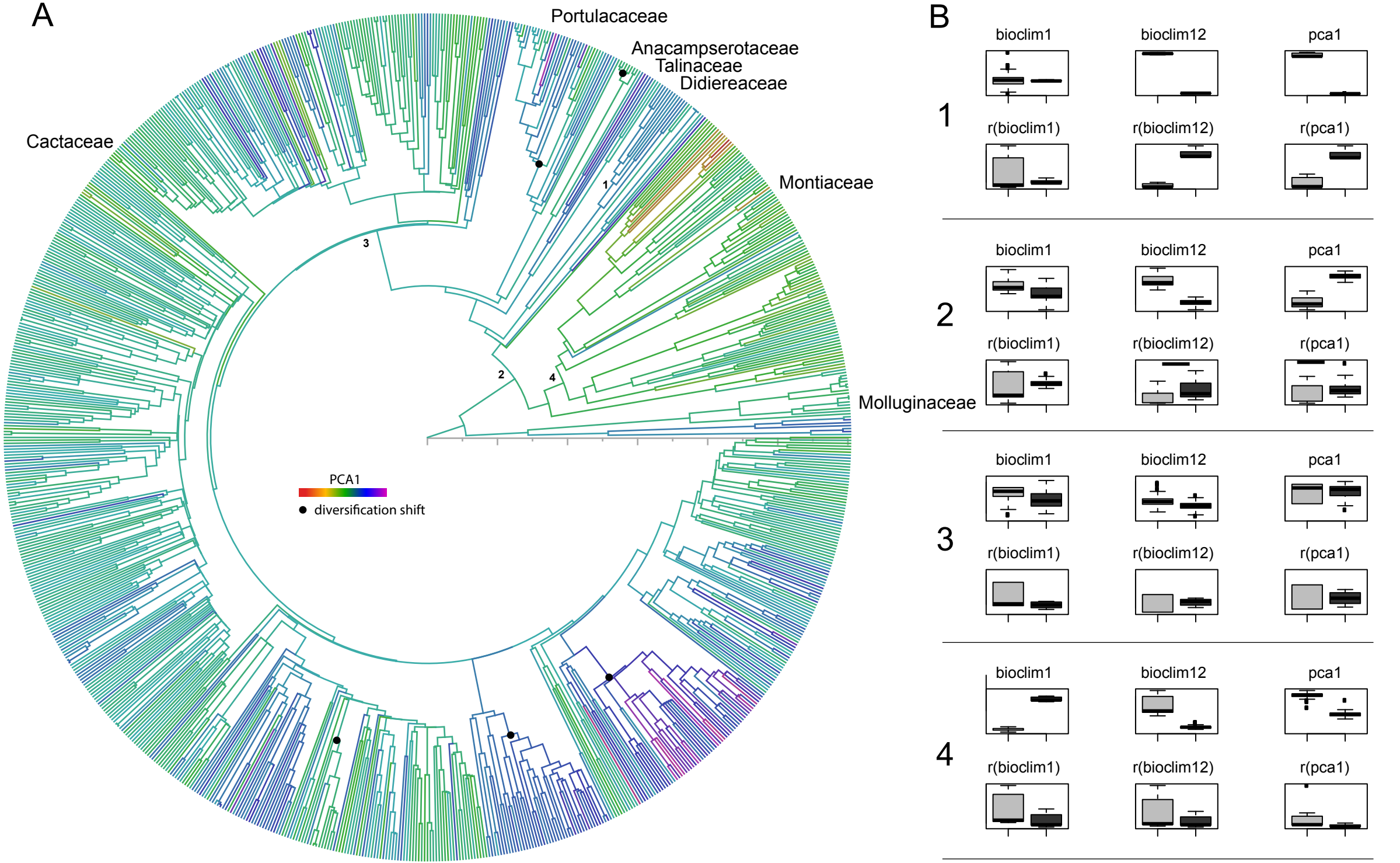
813 NOTE.—TN: tip number; TS: Total lineages under positive selection; Number of lineages under
814 positive selection within certain clades, including Cactaceae (Cac35), Didiereoideae (Did27),
815 Montiaceae (Mont15), and Portulacaceae (Port71). SH80: Duplication nodes with SH-like support >
816 80, Node number can be referred to in figs. 1 and S9. The bolded genes (ABI1 and PP2CA) are
817 important negative regulators in the ABA pathway that mainly discussed in the text. More information
818 about references and gene duplication patterns can be found in table S4.

819
820 Figure Legends
821
822 Fig. 1. Species tree from RAxML analysis of the 841-gene supermatrix. Gene duplication numbers (i.e.,
823 the number of duplication events; number > 12 are shown) are calculated only for branches with strong
824 support (SH-like > 80) in the 8332 rooted clusters.
825
826 Fig. 2. Gene-tree conflict. The pie chart proportions represent genes supporting congruent relationships
827 (concordant), genes supporting the most common conflict (dominant alternative), all other conflicting
828 gene trees (other conflict), and genes with no support (uninformative).
829
830 Fig. 3. Comparison of ln $L$ (sorted by greatest differences in likelihood, $\Delta$ ln$L$) for alternative
831 phylogenetic hypotheses. Left: Three alternative resolutions were compared involving Cactaceae,
832 Portulacaceae, and Anacampserotaceae. Values above the x-axis show the difference in likelihood for
833 those genes supported the resolution found in our result "t1". Those falling below the axis support "t2"
834 (black) or "t3" (gray). Middle: Two alternative hypotheses involving Basellaceae, Didiereaceae and
835 ACPT (Anacampserotaceae, Cactaceae, Portulacaceae, and Talinaceae). Right: Two alternative
836 hypotheses involving the early diverging Cactaceae. Pie chart proportions are the same as in fig. 2.
837 Outlying genes (those genes with extremely high support) are discussed in the Results and Discussion.
838
839 Fig. 4. A. Reconstruction of first axis of principal component analyses (PCA1) on the species level tree
840 of Portulacineae. Black nodes indicate the diversification rate shifts. B. Comparison of MAT (mean
841 annual temperature), MAP (mean annual precipitation), and PCA1 between clades with
842 WGD/extensive gene duplication (1: Didiereoideae, 2: Portulacineae, 3: Cactaceae, 4: Montiaceae
843 except for *Phemeranthus*) and their corresponding sister clades. The upper panel is the comparison of
844 reconstructed state. The lower panel is the comparison of evolutionary rate. Box plots are constructed
845 from bootstrapped phylogenies.


WGD
Ks peak btw 0.05-0.1
4171 # gene dupl.
52/54.5 Support < 97% ML / MQSST (* denotes 100%)
27 Node # referred to in text
Matucana aurantiaca
Echinopsis aurea
Stetsonia coryne
Gymnocalycium mihanovichii
Eriosyce wageneckii
Stenocereus yunkeri
Pachycereus gatesii
Echinocereus pectinatus
Peniocereus cuixmalensis
Rhipsalis baccifera
Copiapoa desertorum
Coryphantha maiz-tablensis
Lophophora williamsii
Ariocarpus retusus
Ferocactus latispinus
Astrophytum myriostigma
Astrophytum asterias
Maihuenia poeppigii
Opuntia cochenillifera
Opuntia bravoana
Opuntia arenaria
Salmiopuntia salmiana
Tunilla corrugata
Tacinga lilae
Tephrocactus bonnieae
Tephrocactus articulatus
Maihueniopsis conoidea
Pterocactus tuberosus
Grusonia bradtiana
Pereskia grandifolia
Pereskia aculeata
Leuenbergeria guamacho
Leuenbergeria bleo
Leuenbergeria lychnidiflora
Cactaceae
Portulaca suffrutescens
Portulaca pilosa
Portulaca amilis
Portulaca grandiflora
Portulaca umbraticola
Portulaca cryptopetala
Portulaca oleracea
Portulaca molokiniensis
Portulacaceae
Anacampseros kurtzii
Anacampseros filamentosa
Talinopsis frutescens
Anacampserotaceae
Talinum paniculatum
Talinum portulacifolium
Talinum fruticosum
Talinaceae
Alluaudia dumosa
Alluaudia humbertii
Alluaudiopsis marnieriana
Alluaudia procera
Didierea trollii
Didierea madagascariensis
Decarya madagascariensis
Portulacaria pygmaea
Portulacaria afra
Didiereaceae
Basella alba
Anredera cordifolia
Basellaceae
Claytonia nevadensis
Claytonia virginica
Montia chamissoi
Lewisia nevadensis
Calandrinia grandiflora
Calyptridium pygmaeum
Calyptridium umbellatum
Cistanthe grandiflora
Phemeranthus parviflorus
Montiaceae
Pharnaceum exiguum
Suessenguthiella caespitosa
Mollugo cerviana
Mollugo nudicaulis
Glinus lotoides
Mollugo verticillata
Mollugo pentaphylla
Molluginaceae
0.04


dominant alternative
other conflict
concordant
uninformative
conflict
concordance
Matucana aurantiaca
Echinopsis aurea
Stetsonia coryne
Gymnocalycium mihanovichii
Eriosyce wagenknechtii
Stenocereus yunckeri
Pachycereus gatesii
Echinocereus pectinatus
Peniocereus cuixmalensis
Rhipsalis baccifera
Copiapoa desertorum
Coryphantha maiz-tablasensis
Lophophora williamsii
Ariocarpus retusus
Ferocactus latispinus
Astrophytum myriostigma
Astrophytum asterias
Maihuenia poeppigii
Opuntia cochenillifera
Opuntia bravoana
Opuntia arenaria
Salmiopuntia salmiana
Tunilla corrugata
Tacinga lilae
Tephrocactus articulatus
Tephrocactus bonnieae
Maihueniopsis conoidea
Pterocactus tuberosus
Grusonia bradtiana
Pereskia aculeata
Pereskia grandifolia
Leuenbergeria bleo
Leuenbergeria guamacho
Leuenbergeria lychnidiflora
Portulaca pilosa
Portulaca suffrutescens
Portulaca amilis
Portulaca grandiflora
Portulaca umbraticola
Portulaca cryptopetala
Portulaca molokiniensis
Portulaca oleracea
Anacampseros filamentosa
Anacampseros kurtzii
Talinopsis frutescens
Talinum paniculatum
Talinum fruticosum
Talinum portulacifolium
Alluaudia humbertii
Alluaudia dumosa
Alluaudia procera
Alluaudiopsis marnieriana
Didierea trollii
Didierea madagascariensis
Decarya madagascariensis
Portulacaria afra
Portulacaria pygmaea
Anredera cordifolia
Basella alba
Claytonia nevadensis
Claytonia virginica
Montia chamissoi
Lewisia nevadensis
Calandrinia grandiflora
Calyptridium pygmaeum
Calyptridium umbellatum
Cistanthe grandiflora
Phemeranthus parviflorus
Pharnaceum exiguum
Suessenguthiella caespitosa
Mollugo cerviana
Mollugo nudicaulis
Glinus dahomensis
Mollugo verticillata
Mollugo pentaphylla
Cactaceae
Portulacaceae
Anacampserotaceae
Talinaceae
Didiereaceae
Basellaceae
Montiaceae
Molluginaceae
Portulacineae
60
50
40
30
20
10
0


A
Cactaceae+relatives
LnL difference
100
50
0
-50
-100
343 genes
t1-t2
t1-t3
t2: 91 genes
t3: 118 genes
genes (sorted)
B
Basellaceae+Didiereaceae
40
20
0
-20
-40
337 genes
185 genes
genes (sorted)
C
early diverging Cactaceae
100
50
0
-50
-100
385 genes
287 genes
genes (sorted)
t1
276/188
Anacampserotaceae
Portulacaceae
Cactaceae
t2
68/397
Portulacaceae
Cactaceae
Anacampserotaceae
t3
85/367
Anacampserotaceae
Cactaceae
Portulacaceae
t1
104/189
ACPT
Didiereaceae
Basellaceae
t2
40/248
Didiereaceae
Basellaceae
ACPT
t1
45/358
core cacti
Pereskia
Leuenbergeria
t2
23/409
Pereskia
Leuenbergeria
core cacti


A
Portulacaceae
Anacampserotaceae
Talinaceae
Didiereaceae
Cactaceae
Montiaceae
Molluginaceae
1
2
3
4
PCA1
diversification shift
B
1
2
3
4
bioclim1
bioclim12
pca1
r(bioclim1)
r(bioclim12)
r(pca1)